1 Introduction

AIMS OF THE BOOK

The primary aim of this book is to provide a compendium of the mol ules that are found at leucocyte surfaces. The book includes entries for 139 mole les. Prior to the entries there are three chapters that provide a perspective for t nking about cell surface molecules. These chapters deal with the dimensions f molecules found at cell surfaces, with the protein superfamilies into whicl cell surface molecules can be grouped, and with the linkages of the genes ncoding the molecules.

A systematic approach to the naming of leucocyte antigens was r ide with the "cluster of differentiation" designation for human leucocyte antig is identified with monoclonal antibodies (mAbs)[1]. MAbs were submitted to w rkshops and were placed into groups based on fluorescent labelling patterns or different leucocyte populations. In this classification, antigens are given a CI number and we provide in this book entries on all the CD antigens. We also incl le entries on antigens from any species which are known to be expressed on t e surface of leucocytes, that have been characterized at the level of their primary equence, but have not been given a CD number. Molecules that have been seque ed and may be at leucocyte surfaces but have not been identified immunochemi ally, are also included.

ORGANIZATION OF THE DATA

Name

The CD nomenclature is used wherever possible for both the humar ntigens and species homologues. In some cases the antigens are known by severa other names and the most common of these are also given together with comm n names for species homologues. The term homologue is reserved for equival t molecules between species and is not used to indicate related molecules in a sup family.

Molecular diagram

This gives a visual representation of the molecule, including ie mode of membrane attachment, the protein domains that are present, and ie degree of glycosylation. Figure 1 lists the symbols that are used to represer the various domains, the N- and O-linked oligosaccharide structures and the dis ohide bonds. A summary of the domains present in the leucocyte molecules is gi n in Table 1 which also includes the page numbers of the individual entries.

For molecules attached to the cell surface by glycosyl-phosphatidy nositol (GPI) anchors, the anchors are shown in the diagrams by an arrow. The c entations of molecules with protein transmembrane sequences are shown by abelling the intracellular terminus.

The criteria used for defining the superfamily relationships of pr ein domains are discussed in Chapter 3. In those cases where no superfamily r ationship is detected, the extracellular region or domain is shown as a circle containing a question mark. The diameter of the circle is proportional to th size of the extracellular region or domain. If the protein sequence contains a hi n proportion of Ser, Thr and Pro residues and is probably heavily O-glycosylated, is shown as an extended structure to distinguish it from regions likely to l ve a folded conformation. It is not possible to show the positions of all the disul ide bonds in

TYPE	SIZE (approximate amino acids)
Complement control protein (CCP)	60
Cytokine receptor (R)	100
Epidermal growth Factor (EGF)	40
Fibronectin type II (FNII)	55
Fibronectin type III (FNIII)	100
Immunoglobulin (Ig) V set	110
Immunoglobulin Ig C1 set	100
Immunoglobulin Ig C2 set	90-100
Lectin C-type	120
Lectin S-type	140
LRG Repeats	24
Link	90

TYPE	SIZE (approximate amino acids)
LDLR	40
Ly-6	70-90
MHC	100
Nerve growth Factor Receptor (NGFR)	40
Scavenger Receptor	110
Somatomedin	40
Tyrosine Kinase	270
Phosphotyrosine Phosphatase (PTPase)	250

OTHER SYMBOLS USED

- N-glycosylation sites
- O-linked glycosylation
- Chondroitin sulphate
- Glycosoaminoglycan
- GPI anchor in Lipid bilayer

Figure 1. *Models for domains and repeats found in leucocyte membrane proteins. These models are used in the diagrams drawn for the entries in Section II. Additional abbreviations used in the domain diagrams are: S, signal sequence; TM, transmembrane sequence; CY, cytoplasmic sequence; G, GPI anchor signal sequence.*

Table 1. *Distribution of domains in leucocyte surface molecules*

Antigen name	Extracellular domains: CCPSF	Cytokine RSF	EGFSF	FN type IISF	FN type IIISF	IgSF	Integrin [a]	Lectin C-typeSF	Lectin S-typeSF	LinkSF	LRG repeat	LDLR	Ly-6SF	MHC	NGFRSF	Scavenger RSF	Somatomedin	Transmembrane regions: Number [b]	GPI [c]	Type [d]	Cytoplasm: Phospho-tyrosine phosphatase	Page number
CD1						1								2				1		1		102
CD2						2												1		1		104
CD3δ						1												1		1		106
CD3ε						1												1		1		106
CD3γ						1												1		1		106
CD4						4												1		1		110
CD5																3		1		1		112
CD6																3		1		1		114
CD7						1												1		1		116
CD8α						1												1		1		118
CD8β						1												1		1		118
CD9																		4		–		120
CD10																		1		2		122
CD11							α											1		1		124
CDw12 (NS)																						129
CD13																		1		2		130
CD14																			1			132
CD15 (CHO)																						135
CD16						2												1 or 1		1		136
CDw17 (CHO)																						139
CD18							β											1		1		140
CD19						2												1		1		142
CD20																		4		1		144
CD21	15 or 16																	1		1		146
CD22						5 or 7												1		1		148
CD23								1										1		2		150
CD24																			1			152
CD25 (IL2Rα)	2																	1		1		154
CD26																		1		2		158
CD27															2			1		1		160
CD28						1												1		1		162
CD29							β											1		1		164
CD30															5			1		1		166
CD31						6												1		1		168
CDw32						2												1		1		170
CD33						2												1		1		174
CD34																		1		1		176
CD35	30																	1		1		178

Antigen name	CCPSF	Cytokine RSF	EGFSF	FN type IISF	FN type IIISF	IgSF	Integrin [a]	Lectin C-typeSF	Lectin S-typeSF	LinkSF	LRG repeat	LDLR	Ly-6SF	MHC	NGFRSF	Scavenger RSF	Somatomedin	Transmembrane regions: Number [b]	Transmembrane regions: GPI [c]	Transmembrane regions: Type [d]	Cytoplasmic: Tyrosine kinase	Cytoplasmic: Phospho-tyrosine phosphatase	Page number
	Extracellular domains																	Transmembrane regions			Cytopla		
CD36																		2?		?			82
CD37																		4					84
CD38																		1		2			85
CD39 (NS)																							87
CD40															4			1		1			88
CD41							α											1		1			90
CD42a											1							1		1			92
CD42bα											7							1		1			92
CD42bβ											1							1		1			92
CD43																		1		1			96
CD44										1								1		1			98
CD45																		1		1		2	02
CD46	4																	1		1			06
CD47 (NS)																							09
CD48						2													1				10
CD49							α											1		1			12
CDw50 (NS)																							17
CD51							α											1		1			18
CD52																			1				20
CD53																		4					22
CD54						5												1		1			24
CD55	4																		1				26
CD56					+	+												1		1			28
CD57 (CHO)																							31
CD58						2												1	α 1	1			32
CD59													1						1				34
CDw60 (CHO)																							37
CD61							β											1		1			38
CD62	9		1					1										1		1			40
CD63																		4					42
CD64						3												1		1			44
CDw65 (CHO)																							47
CD66						4												1		1			48
CD67 (NS)																			1				50
CD68																		1		1			52
CD69 (NS)																							54
CDw70 (NS)																							57
CD71																		1		2			58

Antigen name	CCPSF	Cytokine RSF	EGFSF	FN type IISF	FN type IIISF	IgSF	Integrin [a]	Lectin C-typeSF	Lectin S-typeSF	LinkSF	LRG repeat	LDLR	Ly-6SF	MHC	NGFRSF	Scavenger RSF	Somatomedin	Number [b]	GPI [c]	Type [d]	Tyrosine kinase	Phospho-tyrosine phosphatase	
	Extracellular domains																	Transmembrane regions			Cytoplasm		
CD72								1										1		2			2
CD73																			1				2
CD74																		1		2			2
CD75(CHO)																							2
CD76 (CHO)																							2
CD77 (CHO)																							2
CDw78 (NS)																							2
4-1BB															3			1		1			2
4F2																		1		2			2
114/A10			3															1		1			2
B7						2												1		1			2
B7 (2)																			1				2
B29 (see mIgM)						1												1		1			3
B-G						1												1		1			2
BP-1/6C3																		1		2			2
C5aR																		7					2
c-Kit						5												1		1	1		2
CMRF35						1												1		1			2
CTLA-4						1												1		1			2
Endoglin																		1		2			2
Fas															3			1		1			5
FcαR						2												1		1			2
FcεRIβ																		4					3
FcεRIα						2												1		1			3
FcεRIγ																		1		1			3
flk-2						5												1		1	1		2
fMLPR																		7					3
G-CSFR		1			4	1												1		1			5
GM-CSFRα		1			1													1		1			0
GM-CSFRβ		2			3													1		1			0
gp42						2													1				2
gp49						2												1		1			4
HSA																			1				6
ICAM-2						2												1		1			8
IFNγR																		1		1			0
IL1R						3												1		1			2
IL2Rα (CD25)	2																	1		1			4
IL2Rβ		1			1													1		1			4

Antigen name	CCPSF
IL3Rα	
IL3Rβ	
IL4R	
IL5Rα	
IL5Rβ	
IL6Rα	
IL6Rβ	
IL7R	
IL8R	
LAG-3	
LDLR	
L-Selectin	2
Itk	
Ly-6	
Ly-9	
Ly-49	
Mac-2	
Mannose R	
M-ASGP-BP	
M-CSFR	
MB-1 (see mIgM)	
MDR1	
MHC Class Iα	
MHC Class IIα	
MHC Class IIβ	
MIC2	
mIgM	
MRC OX-2	
MRC OX-40	
MRC OX-47	
NKG2A	
NKR-P1	
PC-1	
R2	
RT6	
Scavenger RI	
Syndecan	
TAPA-1	

Antigen name	Extracellular domains: CCPSF	Cytokine RSF	EGFSF	FN type IISF	FN type IIISF	IgSF	Integrin [a]	Lectin C-typeSF	Lectin S-typeSF	LinkSF	LRG repeat	LDLR	Ly-6SF	MHC	NGFRSF	Scavenger RSF	Somatomedin	Transmembrane regions: Number [b]	GPI [c]	Type [d]	Cytoplasmic: Tyrosine kinase	Tyrosine phosphatase	Page number
IL3Rα		1			1													1		1			326
IL3Rβ		2			2													1		1			326
IL4R		1			1													1		1			330
IL5Rα		1			1													1		1			332
IL5Rβ		2			2													1		1			332
IL6Rα		1			1	1												1		1			334
IL6Rβ		1			4	1												1		1			334
IL7R					1													1		1			338
IL8R																		7					340
LAG-3						4												1		1			342
LDLR			3									7						1		1			344
L-Selectin	2		1					1										1		1			348
Itk																		1		1	1		350
Ly-6													1						1				352
Ly-9						4												1		1			356
Ly-49								1										1		2			358
Mac-2									1									0	0				360
Mannose R				1				8										1		1			362
M-ASGP-BP								1										1		2			364
M-CSFR						5												1		1	1		366
MB-1 (see mIgM)						1												1		1			380
MDR1																		12					368
MHC Class Iα						1								2				1		1			372
MHC Class IIα						1								1				1		1			376
MHC Class IIβ						1								1				1		1			376
MIC2																		1		1			378
mIgM						5												1		1			380
MRC OX-2						2												1		1			384
MRC OX-40																3		1		1			386
MRC OX-47						2												1		2			388
NKG2A								1										1		2			390
NKR-P1								1										1		2			392
PC-1																	2	1		2			394
R2																		4					396
RT6																				1			398
Scavenger RI																	1	1			1		400
Syndecan																		1			1		404
TAPA-1																		4					406

Antigen name	CCPSF	Cytokine RSF	EGFSF	FN type IISF	FN type IIISF	IgSF	Integrin [a]	Lectin C-typeSF	Lectin S-typeSF	LinkSF	LRG repeat	LDLR	Ly-6SF	MHC	NGFRSF	Scavenger RSF	Somatomedin	Number [b]	GPI [c]	Type [d]	Tyrosine kinase	Phospho-tyrosine phosphatase	Page number
	Extracellular domains																	Transmembrane regions			Cytoplasm		
TcR e.g.β						2												1		1			106
Thy-1						1													1				408
TNFRI															4			1		1			410
TNFRII															4			1		1			410

The data are for single polypeptides although some will form homo- or heterodimers.

[a] α or β chain.

[b] The number of predicted passes through the membrane.

[c] Whether or not the molecule has a GPI anchor.

[d] Whether the molecule is a Type I or II transmembrane protein.

NS Sequence not determined.

CHO Determinant dependent on carbohydrate.

the diagrams and these bonds are shown only for Ig-superfamily (IgSF) domains since this has been traditionally done in illustrating this superfamily. The majority of the cytoplasmic domains do not belong to defined superfamilies and are represented by wavy lines whose length is proportional to the sequence length.

The N-linked glycosylation sites are those predicted from the cDNA sequences by the presence of the sequence AsnXaaThr or AsnXaaSer with the exceptions of AsnProThr/Ser, which are normally not glycosylated and AsnXaaSer/ThrPro which are not often glycosylated [2–4]. O-linked glycosylation occurs at Ser or Thr amino acids but there is no clear-cut sequence motif that invariably indicates this type of glycosylation. However, O-linked glycosylation is likely for stretches of sequence that include a preponderance of the amino acids Ser, Thr, Pro [5,6] and this is shown in the diagrams for regions where a high level of O-glycosylation is expected. In many cases there are biochemical data to support the assignment of O-glycosylation.

Glycosaminoglycans (GAG) have been identified in two leucocyte membrane glycoproteins CD44 and syndecan. The number of these sites has not been confirmed biochemically and the symbols indicate only their approximate number and positions. The GAG has been characterized, for example, as chondroitin sulphate in CD44 and this is shown in the diagram.

Size of the processed form of the molecule

The relative molecular weight (M_r) of the polypeptide backbone is calculated from the M_r of the mature protein. This is the polypeptide without the signal sequences for secretion or attachment of a GPI anchor. In most cases the position of cleavage of the signal sequences has been predicted from comparison with consensus sequences from cases where cleavage points are known [7].

The M_r values obtained fro
dodecyl sulphate under reducin
can be considerable variation acc
the value is typical of those in th

The degree of glycosylation

The number of potential N-link
of membrane glycoproteins is
consensus sequence AsnXaaThr
glycoprotein likely to be accou
value of 3000 M_r can be used pe
oligosaccharides can be conside
repeating lactosamine units or c

The number of O-linked sites
thus the following terms are use

(1) Unknown: There are no dat
studies to indica
(2) Nil: No O-linked olig
galactosamine in
apparent M_r on
polypeptide and
(3) Probable +: Sequence data s
(4) + abundant: There are sequer
of O-glycosylatic

Gene location and size

The chromosome location of the
Hughes Medical Institute and
together with its approximate
another species but not for hum

Domain and exon organization

A diagrammatic representation
where it is informative. These
are defined by indicating the p
domain. The first residue is s
terminal side and the second wi
The sequences are given in the
to identify the domain in the f
conserved residues for each dom
alignments given in Chapter 3
each superfamily. Internal posi
full domain since in some cases
of another (e.g. between CD4 d
internal residues should readily
alignments of sequences in Cha

Exons are identified by three
part of the diagram with the an

The M_r values obtained from polyacrylamide gel electrophoresis in sodium dodecyl sulphate under reducing and/or nonreducing conditions are given. There can be considerable variation according to the conditions used for the SDS gels and the value is typical of those in the literature.

The degree of glycosylation

The number of potential N-linked glycosylation sites in the extracellular portions of membrane glycoproteins is shown as determined by the presence of the consensus sequence Asn|XaaThr/Ser. To obtain a rough idea of the weight of the glycoprotein likely to be accounted for by N-linked oligosaccharides an average value of 3000 M_r can be used per structure. It should, however, be noted that some oligosaccharides can be considerably larger due, for example, to the presence of repeating lactosamine units or complex blood group antigens.

The number of O-linked sites cannot be predicted directly from the sequence and thus the following terms are used to describe the extent of O-glycosylation:

(1) Unknown: There are no data from sequence or immunochemical studies to indicate the level of O-glycosylation.

(2) Nil: No O-linked oligosaccharides as determined by the absence of galactosamine in the composition of the molecule or the values for apparent M_r on SDS gels do not exceed those accounted for by the polypeptide and N-linked oligosaccharides.

(3) Probable +: Sequence data suggest the likelihood of O-glycosylation.

(4) + abundant: There are sequence and/or biochemical data to indicate a high level of O-glycosylation.

Gene location and size

The chromosome location of the human gene from the Genome Data Base (Howard Hughes Medical Institute and Johns Hopkins University) or from ref. 8 is given together with its approximate size. If the gene structure has been determined in another species but not for humans, the gene size for the other species is given.

Domain and exon organization

A diagrammatic representation of the positions of domains and exons is given where it is informative. These diagrams are not drawn to scale. Domain positions are defined by indicating the positions of two key conserved residues within the domain. The first residue is shown with two flanking residues on the COOH-terminal side and the second with two flanking residues on the NH_2-terminal side. The sequences are given in the upper part of the domain diagram and can be used to identify the domain in the full sequence given at the end of the entry. The key conserved residues for each domain type are marked with asterisks in the sequence alignments given in Chapter 3 where they can be seen in relation to alignments for each superfamily. Internal positions are identified rather than trying to define the full domain since in some cases the end of one domain overlaps with the beginning of another (e.g. between CD4 domains 1 and 2 [9]). The identification of conserved internal residues should readily allow analysis of domains by comparisons with the alignments of sequences in Chapter 3.

Exons are identified by three or four residues of protein sequence in the lower part of the diagram with the amino acid that is coded at the splice junction aligned

with the intron/exon boundary mark. The type of intron/exon boundary is indicated as type 1 (splicing after nucleotide 1 of the triplet), type 2 (splicing after nucleotide 2) and type 0 (splicing at the end of the codon)[10]. Thus if the junction was type 1 and a Trp residue was shown at the boundary of an exon/intron splice site, the point of junction would be after the T of the TGG codon. By comparing the intron/exon boundaries and the domain designations it should be clear whether a given domain is encoded within one exon. Where exons are known to be alternatively spliced a space is included between the rectangles that represent the exons; all other exons are drawn contiguously.

Tissue distribution
A brief description of the cell types that have been clearly shown to express an antigen is given. In many cases full analyses of tissue distributions have not been done. A failure to mention a cell type cannot be taken as meaning that it is known that the cell type does not express the antigen. The distribution of the CD antigens in humans is reviewed in *Leukocyte Typing IV*[11]. This reference is implied but not cited for each of the CD antigens entries. There are often considerable differences in the tissue distribution between species and this is indicated when data are available. A pullout diagram of the distribution of the human CD antigens on leucocyte populations is enclosed inside the back cover.

Structure
Biochemical and structural data are summarized in this section.

Function
Functional data are summarized in this section.

Comments
This section is used for any additional key points of interest concerning the molecule in question. For example, if the molecule is particularly closely related to another protein or alternative splicing of RNA leads to the production of additional forms of the molecule then this information is included. In addition motifs identified with particular functions are given.

Database entries
The PIR (Protein Identification Resource) and SWISSPROT databases for protein sequences and the GENBANK and EMBL databases for nucleic acid sequences are cited. Every new sequence submitted to the databases is given an accession number. These are then incorporated into entries which are given a specific code. One problem is that codes can change if, for instance, two entries are merged or when sequences become fully annotated. Therefore the only definitive way to identify a sequence is through the accession numbers and these are given in the entries. Where there are multiple accession numbers the primary number is given. In some cases only genomic sequence is available and is spread over several database entries. In these cases more than one accession number is given. For data recently sent to EMBL and GENBANK both databases use the same accession number.

Most software can identify sequences through the accession number and sequences can also be obtained through electronic mail (EMAIL) directly from

automatic computers o
number. The procedures

Obtaining DNA and pro
databases

The EMAIL address is:
The commands used a
specify the EMBL or S
A typical file might lo
GET PROT:P01830
GET NUC:X03152

This file would reques
(protein) and one from

Obtaining protein seque

The EMAIL address is
To retrieve an entry th
database code (not acc
number, type the com
look like:
GET TDRT
ACCESSION A02107

This file would reques
codes that contain the

Obtaining DNA sequen

The GENBANK servic
server at National Cen
address is: retrieve@nc

The file should contai
accession numbers or
DATALIB genbank
BEGIN
J02852

Note for further infor
message to the server
HELP

Sequence
The human amino acid
been obtained from an
amino acids is used (
structural and functio
numbering of the fully
not been defined by seq

automatic computers or servers at the database centres using the acces: ɔn number. The procedures are shown below.

Obtaining DNA and protein sequences from the EMBL and SWISSPROT databases

The EMAIL address is: NETSERV@EMBL-HEIDELBERG.DE
The commands used are: GET to specify the request and NUC or PROT to specify the EMBL or SWISSPROT databases, respectively.
A typical file might look like:
GET PROT:P01830
GET NUC:X03152

This file would request the sending of two entries, one from SWISSPROT (protein) and one from EMBL (DNA).

Obtaining protein sequences from the PIR databases

The EMAIL address is fileserv@gunbrf.bitnet
To retrieve an entry the file should include the command GET together with he database code (not accession number). To obtain the code from the accession number, type the command ACCESSION then the number. A typical file wo .d look like:
GET TDRT
ACCESSION A02107

This file would request the sending of the entry for TDRT and a listing of th codes that contain the accession number A02107

Obtaining DNA sequences from the GENBANK database

The GENBANK service at retrieve@genbank.bio.net has been replaced by a server at National Center of Biotechnology Information (NCBI). The EMAIL address is: retrieve@ncbi.nlm.nih.gov

The file should contain the name of the database, the command BEGIN and ıe accession numbers or key words. A typical file would look like this:
DATALIB genbank
BEGIN
J02852

Note for further information on how to use these servers, send an EMAIL message to the servers containing the single command
HELP

Sequence
The human amino acid sequence is given if it is known. If a sequence has ɔnly been obtained from another species, then this is given. The single letter co . for amino acids is used (defined in Table 1 in Chapter 3). In most cases ı ıere structural and functional data are discussed it is with respect to seq ɔnce numbering of the fully processed form of the molecule. If the NH_2-termini has not been defined by sequence analysis the signal sequence for secretion is pre ɔted

from comparisons with sequences where the position of signal cleavage ha been determined. The consensus rules for defining cleavage sites are reviewed ir ref. 7. The signal sequence is shown on a separate line. For molecules likely t have transmembrane sequences the proposed hydrophobic regions are underlin . For GPI anchors, signal sequences are predicted using criteria discussed in Ch ter 2 and the signal sequence is shown on a separate line below the proposed pr essed sequence. Variants produced by alternative splicing are difficult to repres t but where it is clear that these are expressed the sequences of the alternative fo is are indicated. Allelic alternatives of the gene sequence are given where appropri .

References

It is not feasible to give a comprehensive list of references. The references t at are given are recent ones that should allow access to the rest of the literature a d key references are highlighted in bold. In some cases these are recent reviews ut in others it is the most recent paper containing relevant references. There ar other books in this FactsBook series that specialize on cytokines and their recept s and on adhesion molecules [12,13].

References

1 Bernard, A. et al. (1984) Leukocyte Typing, Springer-Verlag, Berlin, pp.1–8 4.
2 Bause, E. and Hettkamp, H. (1979) FEBS Lett. 108, 341–344.
3 Kornfeld, R. and Kornfeld, S. (1985) Annu. Rev. Biochem. 54, 631–664.
4 Gavel, Y. and von Heijne, G. (1990) Protein Eng. 3, 433–442.
5 Wilson, I.B.H. et al. (1991) Biochem. J. 275, 529–534.
6 Gooley, A.A. et al. (1991) Biochem. Biophys. Res. Commun. 178, 1194–1 1.
7 von Heijne, G. (1983) Eur. J. Biochem. 133, 17–21.
8 McAlpine, P.J. et al. (1990) Cytogenet. Cell Genet. 55, 5–76.
9 Ryu, S.E. et al. (1990) Nature 348, 419–426.
10 Sharp, P.A. (1981) Cell 23, 643–646.
11 Knapp, W. et al. (1989) Leukocyte Typing IV, Oxford University Press, O: ord, UK, pp.1–1182.
12 Callard, R.E. and Gearing, A.J.H. (1993) The Cytokine FactsBook, Acader c Press, London.
13 Pigott, R. and Power, C. (1993) The Adhesion Molecule FactsBook, Acad aic Press, London.

2 Th of

THE DISC ANTIGENS

Early studies

In the 1960s the cell sur obtained fro the human concept of a demonstrate membrane p confirmed b stretch of hy of a fluid m bilayer was by an event surface mo These ideas between me reactions wi of a cell w interpretatic study; for e that do not of considera

The early general me progress w membrane in ref. 8]. I (SDS) woul to the tech allowed re SDS was of destroyed t ionic deter to the hyd with some deoxychola prepared p good solub binding d fractionati proteins b only to hy hydrophob

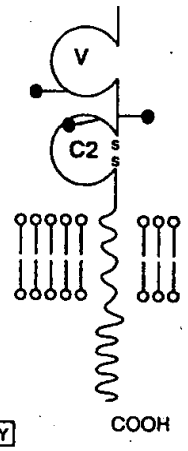
CD2 T11, LFA-2

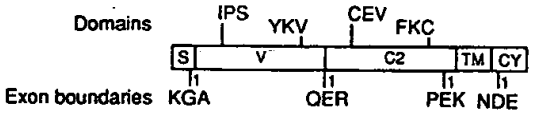
Molecular weights
Polypeptide 36844

SDS PAGE
reduced 45–58 kD
unreduced 45–58 kD

Carbohydrate
N-linked sites 3
O-linked nil

Human gene location and size
1p13; <28.5 kb [1]





Tissue distribution

CD2 is expressed on virtually all T lymphocytes, NK cells and thymocytes [2,3]. In species other than man, CD2 is found on mouse B cells and rat and sheep splenic macrophages [4].

Structure

The extracellular domain is made up of two IgSF domains and is in a subset within the IgSF together with its ligand, CD58 [2]. The structure of the NH_2-terminal domain has been determined by NMR and shown to have an IgSF V-like fold, but lacks the usually conserved disulphide bonds between the β sheets [5]. The NH_2-terminus of the mature polypeptide has been established by protein sequencing[1]. The cytoplasmic domain is rich in basic and proline residues and is well conserved across species [2,4]. The 3′ end of the gene is necessary and sufficient for tissue-specific expression in transgenic mice [1,6].

Function

CD2 interacts through its NH_2-terminal domain with CD58 to form an adhesion pair, for example between T cells and cytotoxic targets or antigen presenting cells [2,3]. Activation of T cells by some pairs of CD2 mAbs is mediated by the cytoplasmic domain and is dependent on the expression of a functional TcR–CD3 complex [2–4]. CD2-mediated signal transduction mimics that through the TcR–CD3 complex [2–4,7]. Activation of NK cells via CD2 is dependent on interaction with CD16 [7].

Comments

The genes for CD2 and CD58 are linked in the genome [2].

Database accession numbers

	PIR	*SWISSPROT*	*EMBL/GENBANK*	*REFERENCE*
Human	A28967	P06729	M16445	*1,8*
Rat	A33071	P08921	X05111	*9*
Mouse	S02293	P08920	Y00023	*10*

Amino acid sequence of human CD2

```
MSFPCKFVAS FLLIFNVSSK GAVS                                        -1
KEITNALETW GALGQDINLD IPSFQMSDDI DDIKWEKTSD KKKIAQFRKE            50
KETFKEKDTY KLFKNGTLKI KHLKTDDQDI YKVSIYDTKG KNVLEKIFDL           100
KIQERVSKPK ISWTCINTTL TCEVMNGTDP ELNLYQDGKH LKLSQRVITH           150
KWTTSLSAKF KCTAGNKVSK ESSVEPVSCP EKGLDIYLII GICGGGSLLM           200
VFVALLVFYI IKRKKQRSRR NDEELETRAH RVATEERGRK PQQIPASTPQ           250
NPATSQHPPP PPGHRSQAPS HRPPPPGHRV QHQPQKRPPA PSGTQVHQQK           300
GPPLPRPRVQ PKPPHGAAEN SLSPSSN                                    327
```

References

1 **Moingeon, P. et al. (1989) Immunol. Rev. 111, 111–144.**
2 **Beyers, A.D. et al. (1989) Immunol Rev. 111, 59–77.**
3 **Bierer, B.E. and Burakoff, S.J.(1989) Immunol. Rev. 111, 267–294.**
4 Driscoll, P.C. et al. (1991) Nature 353, 762–765.
5 Lang, G. et al. (1988) EMBO J. 7,1675–1682.
6 Greaves, D.R. et al. (1989) Cell 56, 979–986.
7 Spruyt, L.L. et al. (1991) J. Exp Med. 174,1407–1415.
8 Seed, B. and Aruffo, A. (1987) Proc. Natl Acad. Sci. USA 84, 3365–3369.
9 Sewell,W.A. et al.(1987) Eur. J. Immunol. 17,1015–1020.
10 Williams, A.F. et al.(1987) J. Exp. Med.165, 368–380.

CD4

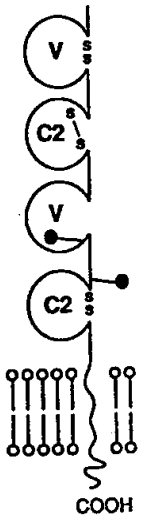
T4, L3T4 (mouse), W3/25 (rat)

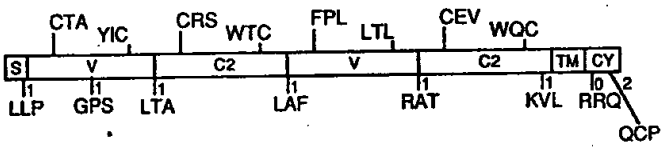
Molecular weights
Polypeptide 48400

SDS PAGE
reduced 55 kD
unreduced 55 kD

Carbohydrate
N-linked sites 2
O-linked nil

Human gene location and size
12pter-p12; 33 kb [1]





Tissue distribution

CD4 is expressed on most thymocytes and approximately two thirds of peripheral blood T cells, which constitute the CD8 negative cells [2]. In human and rat but not in mouse, CD4 is expressed on monocytes and macrophages [2].

Structure

The extracellular domain is made up of four IgSF domains. The structures of the amino terminal two domains have been determined by X-ray crystallography, confirming that they are Ig-like [3,4]. Domain 2 is characterized by an unusual disulphide within one beta sheet and domain 3 lacks a disulphide in the position conserved in most IgSF domains. Cat CD4 shows some unusual features with 17 residues inserted between domains 1 and 2 [5]. There is an additional Cys in domain 1 and the Cys in the unusual β strand C position in domain 2 is replaced with a Trp and there is an extra Cys in the β strand F [5]. The position of the NH_2-terminus has been established for the rat homologue [6].

Function

CD4 is an accessory molecule in the recognition of foreign antigens in association with MHC Class II antigens by T cells [2]. MAbs against CD4

inhibit T cell fu
CD4 is phosph
cells are activat
interacts with a
motif shown be
binding of the
domain [3,4].

Comments

CD4 shows par
protein (see pag

Motifs involv

p56[lck] recogniti

RRQAERMSQI K

Database acces

Human
Rat
Mouse

Amino acid s

MNRGVPFRHL
KKVVLGKKGD
LNDRADSRRS
TANSDTHLLQ
LQDSGTWTCT
AFTVEKLTGS
KKLPLHLTLP
LTCEVWGPTS
GQVLLESNIK
RRQAERMSQI

References

1 Maddon, P
2 Parnes, J.R
3 Wang, J. e
4 Ryu, S-E.
5 Norimine
6 Clark, S.J.
7 Shin, J. et
8 Turner, J.

inhibit T cell functions *in vivo* and *in vitro* [2]. The cytoplasmic domain of CD4 is phosphorylated at Ser residues 408, 415, 431 (see below) when T cells are activated by antigen or phorbol esters [7]. The cytoplasmic domain interacts with a lymphocyte-specific tyrosine kinase called $p56^{lck}$ through a motif shown below [8]. CD4 is a receptor for HIV-1 (AIDS virus) and the binding of the viral gp120 protein is to a region of the amino terminal domain [3,4].

Comments

CD4 shows particularly close similarities in overall structure to the LAG-3 protein [see page 342].

Motifs involved in CD4 function

$p56^{lck}$ recognition site (underlined) and Ser residues phosphorylated (in bold)

RRQAERMSQI KRLLSEKKTC QCPHRFQKTC SPI (433)

Database accession numbers

	PIR	*SWISSPROT*	*EMBL/GENBANK*	*REFERENCE*
Human	AO2109	P01730	M12807	2
Rat	A27449	P05540	M15768	6
Mouse	A02110	P06332	M13816	2

Amino acid sequence of human CD4

```
MNRGVPFRHL LLVLQLALLP AATQG                                    -1
KKVVLGKKGD TVELTCTASQ KKSIQFHWKN SNQIKILGNQ GSFLTKGPSK        50
LNDRADSRRS LWDQGNFPLI IKNLKIEDSD TYICEVEDQK EEVQLLVFGL       100
TANSDTHLLQ GQSLTLTLES PPGSSPSVQC RSPRGKNIQG GKTLSVSQLE       150
LQDSGTWTCT VLQNQKKVEF KIDIVVLAFQ KASSIVYKKE GEQVEFSFPL       200
AFTVEKLTGS GELWWQAERA SSSKSWITFD LKNKEVSVKR VTQDPKLQMG       250
KKLPLHLTLP QALPQYAGSG NLTLALEAKT GKLHQEVNLV VMRATQLQKN       300
LTCEVWGPTS PKLMLSLKLE NKEAKVSKRE KAVWVLNPEA GMWQCLLSDS       350
GQVLLESNIK VLPTWSTPVQ PMALIVLGGV AGLLLFIGLG IFFCVRCRHR       400
RRQAERMSQI KRLLSEKKTC QCPHRFQKTC SPI                         433
```


References

1 Maddon, P.J. et al. (1987) Proc. Nat Acad. Sci. USA 84, 9155–9159.
2 **Parnes, J.R. (1989) Adv. Immunol. 44, 265–311.**
3 Wang, J. et al. (1990) Nature 348, 411–418.
4 Ryu, S-E. et al. (1990) Nature 348, 419–426.
5 Norimine, J. et al. (1992) Immunology 75, 74–79.
6 Clark, S.J. et al. (1987) Proc. Natl Acad. Sci. USA 84, 1649–1653.
7 Shin, J. et al. (1990) EMBO J. 9, 425–434.
8 Turner, J.M. et al. (1990) Cell 60, 755–757.

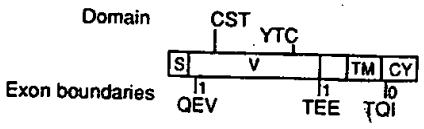
CD7 gp40

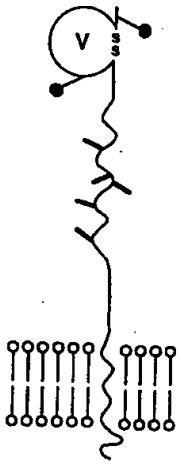
Molecular weights
Polypeptide 22919

SDS PAGE
reduced 40 kD
unreduced 38 kD

Carbohydrate
N-linked sites 2
O-linked probable +

Human gene size and location
17; 3 kb[1]





Tissue distribution

CD7 is the earliest marker antigen expressed in the T lineage being found on T cell precursors in fetal liver and thorax prior to thymic colonization and in thymus and bone marrow [2]. CD7 is expressed on pluripotential haematopoietic cells, most human thymocytes and a major subset of peripheral blood T cells [2]. CD7 is a marker for pluripotential stem cell leukaemias and T cell acute lymphocytic leukaemia [2].

Structure

An Ig-like domain is separated from the transmembrane sequence by 4 repeats containing a high proportion of Pro, Ser and Thr residues [3]. This region is likely to be O-glycosylated and to form an extended structure like that proposed for the CD8 hinge. The gene for CD7 is similar to the murine *Thy-1* gene which places it in a class of tissue-specific genes whose promoters lack a TATA element [1].

Function

The function of CD7 is unknown and no ligand has yet been identified. A proposal that CD7 was an IgM receptor was not confirmed in studies on expression from cDNA [3]. A role for CD7 in the activation of T cells with γ/δ receptors has been proposed based on mAb-induced activation [3].

Database accession numbers

	PIR	*SWISSPROT*	*EMBL/GENBANK*	*REFERENCE*
Human	S03520	P09564	X06180	3



Amino acid sequence of human CD7

```
MAGPPRLLLL PLLLALARGL PGALA                                    -1
AQEVQQSPHC TTVPVGASVN ITCSTSGGLR GIYLRQLGPQ PQDIIYYEDG          50
VVPTTDRRFR GRIDFSGSQD NLTITMHRLQ LSDTGTYTCQ AITEVNVYGS         100
GTLVLVTEEQ SQGWHRCSDA PPRASALPAP PTGSALPDPQ TASALPDPPA         150
ASALPAALAV ISFLLGLGLG VACVLARTQI KKLCSWRDKN SAACVVYEDM         200
SHSRCNTLSS PNQYQ                                               215
```



References

1 Schanberg, L.E. et al. (1991) Proc. Natl Acad. Sci. USA. 88, 603–607.
2 **Haynes, B.F. et al. (1989) Immunol. Today 10, 87–91.**
3 **Aruffo, A. and Seed, B. (1987) EMBO J. 6, 3313–3316.**
4 Carrel, S. et al. (1991) Eur. J. Immunol. 21, 1195–1200.

CD16 FcγRIII

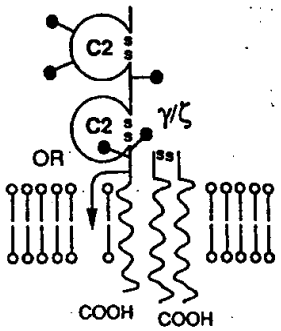
Molecular weights

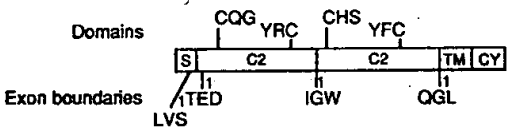
Polypeptide	Transmembrane form	27268
SDS PAGE reduced		50–80 kD

Carbohydrate

N-linked sites	5
O-linked	unknown

Human gene location and size
1q23; 9 kb [1,2]





Tissue distribution

In humans the transmembrane form of CD16 is expressed on NK cells and macrophages, whilst the GPI-linked form is expressed on neutrophils. In the mouse no GPI-linked form of CD16 has been identified and the transmembrane form is expressed on macrophages, NK cells, neutrophils and myeloid precursors [reviewed in ref. 3].

Structure

There are two distinct forms of human CD16 encoded by two linked genes: a transmembrane form with a 25 amino acid cytoplasmic tail and a glycosyl-phosphatidylinositol (GPI)-linked form [1,2,4]. Their extracellular sequences differ by only six amino acids and site directed mutagenesis showed that amino acid Ser_{186} (mature numbering) determines the attachment of the GPI anchor [3,5]. The extracellular region of CD16 is formed by two C2-set IgSF domains [4]. The transmembrane form is non-covalently associated with the FcεRIγ chain [5] or the TcRζ chain [6], forming a multimeric complex, and its surface expression is dependent on the co-expression of the γ or ζ chains [5,6]. It has been shown that on mast cells CD16 is also associated with the β chain of the FcεRI [7]. There are two alleles of the GPI-linked form [3]. Both CD16 genes have identical intron exon boundaries [3].

Function

CD16 is a low affinity receptor for aggregated IgG [3]. The transmembrane form binds IgG complexed to antigens and mediates phagocytosis and antibody-dependent cellular cytotoxicity. Signal transduction by CD16 is mediated through the γ chain [8]. Cross-linking of CD16 on NK cells with immune complexes or mAbs induces calcium mobilization and hydrolysis of membrane phosphoinositides. In contrast the GPI-linked form of neutrophils binds the ligands but is unable to induce any signal or

functional effect [reviewed in ref. 3]. The engagement of the transmembrane form results in the tyrosine phosphorylation of the TcRζ chain [9].

Comments

CD16 (FcγRIII) has structural similarity to CD64 (FcγRI), CDw32 (FcγRII) and FcεRIα [3]. The sequence LFAVDTGL is completely conserved in the transmembrane regions of human and mouse CD16 and human, mouse and rat FcεRIα [10]. A family of CD16 isoforms have been described in the rat [10]. Patients with paroxysmal nocturnal haemoglobinuria lack the GPI-linked form of CD16 and this is probably responsible for the presence of circulating immune complexes and the susceptibility to bacterial infections associated with this disease [11].

Database accession numbers

	PIR	*SWISSPROT*	*EMBL/GENBANK*	*REFERENCE*
Human	JL0107	P08637	X16863	*1*
Mouse		P08508	M14215	*12*
Rat			M64368-M64370	*10*

Amino acid sequence of human CD16

Transmembrane form

```
MWQLLLPTAL LLLVSAG                                          -1
MRTEDLPKAV VFLEPQWYRV LEKDSVTLKC QGAYSPEDNS TQWFHNESLI      50
SSQASSYFID AATVDDSGEY RCQTNLSTLS DPVQLEVHIG WLLLQAPRWV     100
FKEEDPIHLR CHSWKNTALH KVTYLQNGKG RKYFHHNSDF YIPKATLKDS     150
GSYFCRGLFG SKNVSSETVN ITITQGLAVS TISSFFPPGY QVSFCLVMVL     200
LFAVDTGLYF SVKTNIRSST RDWKDHKFKW RKDPQDK                   237
```

GPI-linked form

```
MWQLLLPTAL LLLVSAG                                          -1
MRTEDLPKAV VFLEPQWYRV LEKDSVTLKC QGAYSPEDNS TQWFHNESLI      50
SSQASSYFID AATVDDSGEY RCQTNLSTLS DPVQLEVHIG WLLLQAPRWV     100
FKEEDPIHLR CHSWKNTALH KVTYLQNGKG RKYFHHNSDF YIPKATLKDS     150
GSYFCRGLFG SKNVSSETVN ITITQGLAVS TISSFS                    186
PPGYQVSFCL VMVLLFAVDT GLYFSVKTNI                           216
```


References

1 Ravetch, J.V. and Perussia, B. (1989) J. Exp. Med. 170, 481–497.
2 Qiu, W.Q. et al. (1990) Science 248, 732–735.
3 **Ravetch, J.V. and Kinet, J.-P. (1991) Annu. Rev. Immunol. 9, 457–492.**
4 Simmons, D. and Seed, B. (1988) Nature 333, 568–570.
5 Hibbs, M.L. et al. (1989) Science 246, 1608–1611.
6 Lanier, L.L. et al. (1989) Nature 342, 803–806.
7 Kurosaki, T. et al. (1992) J. Exp. Med. 175, 447–451.
8 Wirthmueller, U. et al. (1992) J. Exp. Med. 175, 1381–1390.
9 O'Shea, J.J. et al. (1991) Proc. Natl Acad. Sci. USA 88, 350–354.
10 Farber, D.L. and Sears, D.W. (1991) J. Immunol. 146, 4353–4361.
11 Selvaraj, P. et al. (1988) Nature 333, 565–567.
12 Ravetch, J.V. et al. (1986) Science 234, 718–725.

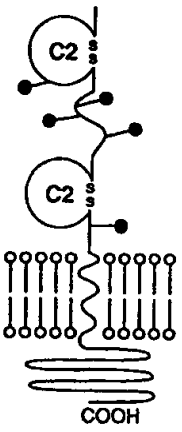
CD19 B4

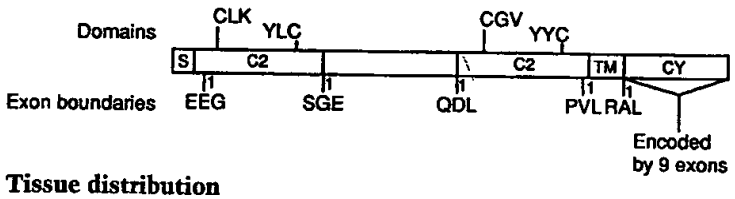
Molecular weights
Polypeptide 59154

SDS PAGE
reduced 95 kD

Carbohydrate
N-linked sites 5
O-linked unknown

Human gene size
8 kb [1]





Tissue distribution

CD19 is expressed on all human B cells and on B cell precursors before CD10 (CALLA antigen) and the intracytoplasmic μ chain. It is not expressed on plasma cells. CD19 is also present on follicular dendritic cells [2]. Mouse CD19 mRNA is expressed in B cell lines and spleen [1].

Structure

The extracellular region of CD19 consists of two C2-set IgSF domains separated by a region that contains a pair of Cys residues but has no significant sequence homology with any known protein [2]. The large cytoplasmic domain is highly conserved between species [1] and contains several potential phosphorylation sites on Ser and Thr and nine Tyr residues [2].

Function

CD19 is involved in the regulation of B cell proliferation. Coligation of CD19 with mIg decreases the threshold for mIg-dependant stimulation [3]. However cross-linking of CD19 molecules independently from mIg inhibits the increase in intracellular calcium and proliferation induced by anti-Ig [4]. In addition, antibody cross-linking of CD19 alone in B and pre-B cells induces an increase in the intracellular calcium levels, without leading to B cell activation [4,5]. CD19 associates with the complement receptor 2 (CD21) [6] TAPA-1 and Leu13 antigens [7] forming a signal transduction complex on the membrane of B cells.

Database accessio

Human
Mouse
Guinea pig

Amino acid seq

MPPPRLLFFL LF
EVRPEEPLVV KV
GLPGLGIHMR PL
GSGELFRWNV SD
WEGEPPCVPP RD
VHPKGPKSLL SL
TMSFHLEITA RP
LVLRRKRKRM TD
WAAGLGGTAP SY
DSEFYENDSN LG
ELTQPVARTM DF
RGQPGPNHEE DA

References

1 Zhou, L.-J. et
2 **Tedder, T.F. a**
3 **Carter, R.H. a**
4 Pezzuto, A. et
5 Ledbetter, J.A
6 **Matsumoto, A**
7 Bradbury, L. e

Database accession numbers

	PIR	SWISSPROT	EMBL/GENBANK	REFERENCE
Human			M28170	2
Mouse			M62542	1
Guinea pig			M62543	1

Amino acid sequence of human CD19

```
MPPPRLLFFL LFLTPM                                                  -1
EVRPEEPLVV KVEEGDNAVL QCLKGTSDGP TQQLTWSRES PLKPFLKLSL             50
GLPGLGIHMR PLASWLFIFN VSQQMGGFYL CQPGPPSEKA WQPGWTVNVE            100
GSGELFRWNV SDLGGLGCGL KNRSSEGPSS PSGKLMSPKL YVWAKDRPEI            150
WEGEPPCVPP RDSLNQSLSQ DLTMAPGSTL WLSCGVPPDS VSRGPLSWTH            200
VHPKGPKSLL SLELKDDRPA RDMWVMETGL LLPRATAQDA GKYYCHRGNL            250
TMSFHLEITA RPVLWHWLLR TGGWKVSAVT LAYLIFCLCS LVGILHLQRA            300
LVLRRKRKRM TDPTRRFFKV TPPPGSGPQN QYGNVLSLPT PTSGLGRAQR            350
WAAGLGGTAP SYGNPSSDVQ ADGALGSRSP PGVGPEEEEG EGYEEPDSEE            400
DSEFYENDSN LGQDQLSQDG SGYENPEDEP LGPEDEDSFS NAESYENEDE            450
ELTQPVARTM DFLSPHGSAW DPSREATSLG SQSYEDMRGI LYAAPQLHSI            500
RGQPGPNHEE DADSYENMDN PDGPDPAWGG GGRMGTWSTR                       540
```

References

1 Zhou, L.-J. et al. (1991) J. Immunol. 147, 1424–1432.
2 **Tedder, T.F. and Isaacs, C.M. (1989) J. Immunol. 143, 712–717.**
3 **Carter, R.H. and Fearon, D.T. (1992) Science 256, 105–107.**
4 Pezzuto, A. et al. (1987) J. Immunol. 138, 2793–2799.
5 Ledbetter, J.A. et al. (1988) Proc. Natl Acad. Sci. USA 85, 1897–1901.
6 **Matsumoto, A.K. et al. (1991) J. Exp. Med. 173, 55–64.**
7 Bradbury, L. et al. (1991) FASEB J. 5, A611.

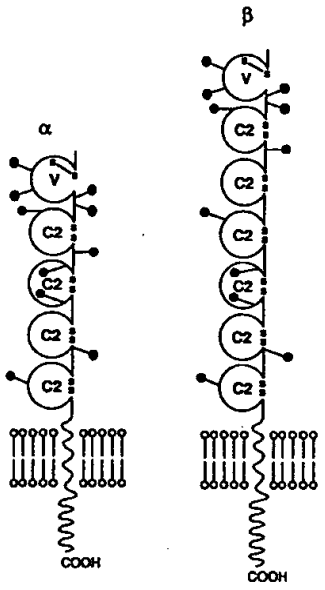
CD22 BL-CAM

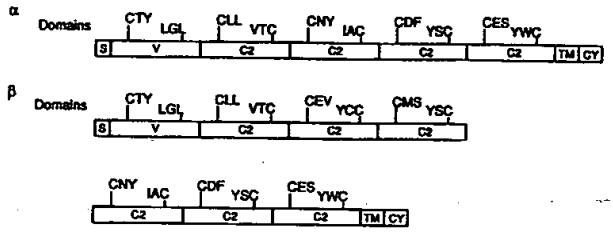
Molecular weights

Polypeptide	CD22α 70991
SDS PAGE	
reduced	130 kD/140 kD heterodimer
unreduced	120 kD/130 kD heterodimer

Carbohydrate

N-linked sites	CD22α, 10 CD22β, 11
O-linked	CD22α, nil CD22β, nil





Tissue distribution

CD22 is expressed on a subset of Ig$^+$ mature B cells (60–80%) and on hairy cell leukaemia cells. It is also present in the cytoplasm of pro-B, pre-B and mature B cells. CD22 expression is lost after B cell activation [1].

Structure

Two forms of CD22 have been described. The α form of CD22 has five IgSF domains [1]. The β form is identical to the α form except that it has two additional IgSF domains (domains three and four) and a cytoplasmic domain 23 amino acids longer than that of the α form [2,3]. It has not been established whether these forms arise by alternative splicing. The sequence of CD22 is similar to that of myelin-associated glycoprotein [1,2] and both have been predicted to have an unusual disulphide bond between β strands B and E in the first domain [4]. The cytoplasmic region contains possible protein kinase C and casein kinase II phosphorylation sites and six Tyr residues [2].

Function

CD22α has been shown to mediate adhesion of transfected cells to monocytes and erythrocytes [1] and CD22β to mediate adhesion of B cells and CD4+ T cells [2,3]. The CD45RO antigen on T cells and the CD75 antigen on B cells have been identified as ligands for CD22β [2]. CD22+ cells, but not CD22- cells, respond to cross-linking with anti-Ig with an increase in intracellular calcium and a strong proliferative response [5]. MAb cross-linking of CD22 antigen augments the increase in calcium and the proliferation induced by anti-Ig [5].

Comments

CD22 is an heterodimer of 130 and 140 kD glycoproteins that originate from two different precursor molecules [6].

Database accession numbers

	PIR	*SWISSPROT*	*EMBL/GENBANK*	*REFERENCE*
Human	A35648	P20273		1

Amino acid sequence of human CD22α

```
MHLLGPWLLL LVLEYLAFS                                          -1
DSSKWVFEHP ETLYAWEGAC VWIPCTYRAL DGDLESFILF HNPEYNKNTS        50
KFDGTRLYES TKDGKVPSEQ KRVQFLGDKN KNCTLSIHPV HLNDSGQLGL       100
RMESKTEKWM ERIHLNVSER PFPPHIQLPP EIQESQEVTL TCLLNFSCYG       150
YPIQLQWLLE GVPMRQAAVT STSLTIKSVF TRSELKFSPQ WSHHGKIVTC       200
QLQDADGKFL SNDTVQLNVK HPPKKVTTVI QNPMPIREGD TVTLSCNYNS       250
SNPSVTRYEW KPHGAWEEPS LGVLKIQNVG WDNTTIACAA CNSWCSWASP-      300
VALNVQYAPR DVRVRKIKPL SEIHSGNSVS LQCDFSSSHP KEVQFFWEKN       350
GRLLGKESQL NFDSISPEDA GSYSCWVNNS IGQTASKAWT LEVLYAPRRL       400
RVSMSPGDQV MEGKSATLTC ESDANPPVSH YTWFDWNNQS LPYHSQKLRL       450
EPVKVQHSGA YWCQGTNSVG KGRSPLSTLT VYYSPETIGR RVAVGLGSCL       500
AILILAICGL KLQRRWKRTQ SQQGLQENSS GQSFFVRNKK VRRAPLSEGP       550
HSLGCYNPMM EDGISYTTLR FPEMNIPRTG DAESSEMQRP PPDCDDTVTY       600
SALHKRQVGT MRTSFQIFQK MRGFITQS                               628
```

Amino acid sequence of human CD22β

Sequence not available in EMBL/GENBANK databases.

References

1 Stamenkovic, I. and Seed, B. (1990) Nature 345, 74–77.
2 Wilson, G.L. et al. (1991) J. Exp. Med. 173, 137–146.
3 **Stamenkovic, I. et al. (1991) Cell 66, 1133–1144.**
4 Williams, A.F. et al. (1989) Cold Spring Harbor Symp. Quant. Biol. LIV, CSH Press, pp.637–647.
5 Pezzutto, A. et al. (1988) J. Immunol. 140, 1791–1795.
6 Schwartz-Albiez, R. et al. (1991) Int. Immunol. 3, 623–633.

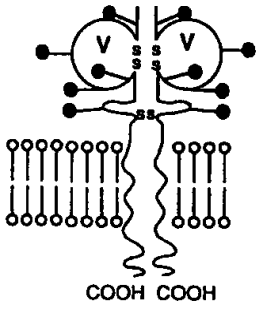
CD28 Tp44

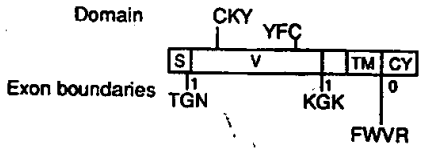
Molecular weights
Polypeptide 23085

SDS PAGE
reduced 44 kD
unreduced 90 kD

Carbohydrate
N-linked sites 5
O-linked unknown

Human gene location and size
2q33–q34; 36 kb [1]





Tissue distribution

CD28 is expressed on most T lineage cells and plasma cells [2]. Mature thymocytes have higher levels of CD28 than the immature cells and among peripheral T cells, 95% of $CD4^+$ cells and 50% of $CD8^+$ cells are positive [2]. Activation of T cells leads to enhanced CD28 expression [2].

Structure

CD28 is a member of the IgSF and is expressed as a disulphide-linked homodimer [2,3]. Human and mouse CD28 are 68% identical at the amino acid level [4]. CD28 is particularly similar to CTLA-4 with which it shares a ligand and probably a common ancestor in evolution [4].

Function

The ligand for CD28 is B7 [5,6] which is expressed on activated B cells, suggesting an important role for CD28 in the interaction between T and B cells. Activation of T cells via CD28 has provided evidence for a CD28 signalling pathway which involves stabilization of cytokine mRNA levels and is separate from that used by the TcR–CD3 complex [2,7].

Database accession numbers

	PIR	*SWISSPROT*	*EMBL/GENBANK*	*REFERENCE*
Human		P10747	J02988	*3*
Mouse			M34563	*4*
Rat			X55288	*8*



Amino acid sequence of human CD28

```
MLRLLLALNL FPSIQVTG                                              -1
NKILVKQSPM LVAYDNAVNL SCKYSYNLFS REFRASLHKG LDSAVEVCVV            50
YGNYSQQLQV YSKTGFNCDG KLGNESVTFY LQNLYVNQTD IYFCKIEVMY           100
PPPYLDNEKS NGTIIHVKGK HLCPSPLFPG PSKPFWVLVV VGGVLACYSL           150
LVTVAFIIFW VRSKRSRLLH SDYMNMTPRR PGPTRKHYQP YAPPRDFAAY           200
RS                                                               202
```



References

1 Lee, K.P. et al. (1990) J. Immunol. 145, 344–352.
2 **June, C.H. et al. (1990) Immunol. Today 11, 211–216.**
3 Aruffo, A. and Seed, B. (1987) Proc. Natl Acad. Sci. USA 84, 8573–8577.
4 Gross, J.A. et al. (1990) J. Immunol.144, 3201–3210.
5 Linsley, P.S. et al. (1990) Proc. Natl Acad. Sci. USA 87, 5031–5035.
6 Linsley, P.S. et al. (1991) J. Exp. Med. 173, 721–730.
7 Lindsten, T. et al. (1989) Science 244, 339–42.
8 Clark, G.S and Dallman, M.J. (1992) Immunogenetics 35, 54–57.

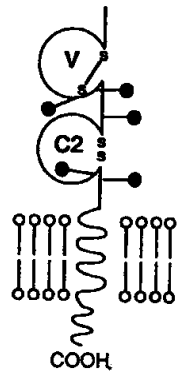
CD33

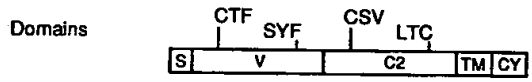
Molecular weights
Polypeptide 37908

SDS PAGE
reduced 67 kD

Carbohydrate
N-linked sites 5
O-linked nil

Human gene location and size
19 q 13.3; <35 kb [1]





Tissue distribution

CD33 is found on granulocyte and macrophage precursors in the bone marrow, but is not on pluripotent stem cells. The protein is also expressed on peripheral monocytes, and is a useful marker for distinguishing myelogenous leukaemia cells from lymphoid or erythroid leukaemias [2].

Structure

The cDNA sequence of CD33 codes for two IgSF domains with closest similarity to the NH_2-terminal domains of myelin-associated glycoprotein and CD22 [2]. The disulphide bond in domain 1 of CD33 is predicted to be between β strands B and E with a Phe residue in the position of the conserved Cys of β strand F [3]. There is likely to be a disulphide bond between domains 1 and 2.

Function

Unknown.

Comments

Expressed by myeloid precursors after CD34 [4]. The CD33 gene is located near the gene for myelin-associated glycoprotein which has been localized to position 19 q 13.1.

Database accession numbers

	PIR	*SWISSPROT*	*EMBL/GENBANK*	*REFERENCE*
Human	A30521	P20138	M23197	2

Amino acid s

MPLLLLLPLL
DPNFWLQVQE
SGDSPVATNK
FRMERGSTKY
ACEQGTPPIF
GAGVTTERTI
TALLALCLCL
GPTETSSCSG

References

1 Peiper, S.C.
2 Simmons, D
3 Williams, A
Quantitativ
4 Andrews, R



Amino acid sequence of human CD33

```
MPLLLLLPLL WAGALAM                                          -1
DPNFWLQVQE SVTVQEGLCV LVPCTFFHPI PYYDKNSPVH GYWFREGAII      50
SGDSPVATNK LDQEVQEETQ GRFRLLGDPS RNNCSLSIVD ARRRDNGSYF     100
FRMERGSTKY SYKSPQLSVH VTDLTHRPKI LIPGTLEPGH SKNLTCSVSW     150
ACEQGTPPIF SWLSAAPTSL GPRTTHSSVL IITPRPQDHG TNLTCQVKFA     200
GAGVTTERTI QLNVTYVPQN PTTGIFPGDG SGKQETRAGV VHGAIGGAGV     250
TALLALCLCL IFFIVKTHRR KAARTAVGRN DTHPTTGSAS PKHQKKSKLH     300
GPTETSSCSG AAPTVEMDEE LHYASLNFHG MNPSKDTSTE YSEVRTQ        347
```


References

1 Peiper, S.C. et al. (1988) Blood 72, 314–321.

2 **Simmons, D. and Seed, B. (1988) J. Immunol. 141, 2797–2800.**

3 Williams, A.F. et al. (1989) Cold Spring Harbor Symposia on Quantitative Biology 54, 637–647.

4 Andrews, R.G. et al. (1989) J. Exp. Med. 169, 1721–1731.

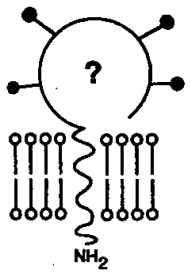
CD38 T10

Molecular weights
Polypeptide 34302

SDS PAGE
reduced 45 kD
unreduced 45 kD

Carbohydrate
N-linked sites 4
O-linked unknown

Human gene location
4



Tissue distribution

CD38 is found on early cells of the B and T cell lineages, on activated B and T cells, but not on most mature resting peripheral lymphocytes. It is present on thymocytes, pre-B cells, germinal centre B cells, mitogen-activated T cells, and Ig-secreting plasma cells [1,2].

Structure

CD38 is a type II membrane glycoprotein, with the transmembrane sequence near the NH_2-terminus [1].

Function

Unknown, but anti-CD38 mAbs in combination with accessory cells are mitogenic for T cells and a variety of cell lines [3].

Database accession numbers

	PIR	*SWISSPROT*	*EMBL/GENBANK*	*REFERENCE*
Human			M34461	1

Amino acid sequence of human CD38

```
MANCEFSPVS GDKPCCRLSR RAQLCLGVSI LVLILVVVLA VVVPRWRQTW   50
SGPGTTKRFP ETVLARCVKY TEIHPEMRHV DCQSVWDAFK GAFISKHPCN  100
ITEEDYQPLM KLGTQTVPCN KILLWSRIKD LAHQFTQVQR DMFTLEDTLL  150
GYLADDLTWC GEFNTSKINY QSCPDWRKDC SNNPVSVFWK TVSRRFAEAA  200
CDVVHVMLNG SRSKIFDKNS TFGSVEVHNL QPEKVQTLEA WVIHGGREDS  250
RDLCQDPTIK ELESIISKRN IQFSCKNIYR PDKFLQCVKN PEDSSCTSEI  300
```

References

1 **Jackson, D.G. and Bell, J. I. (1990) J. Immunol. 144, 2811–2815.**
2 Alessio, M. et al. (1990) J. Immunol. 145, 878–884.
3 Funaro, A. et al. (1990) J. Immunol. 145, 2390–2396.

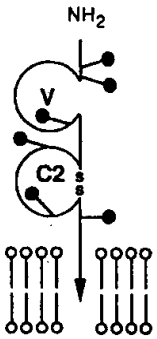
CD48 Blast-1, HuLy-m3, BCM1 antigen, MRC OX-45

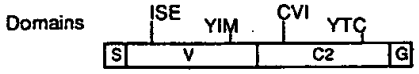
Molecular weights
Polypeptide 24854

SDS PAGE
reduced 40–47 kD

Carbohydrate
N-linked sites 6
O-linked unknown

Human gene location
1q21–q23





Tissue distribution

Expressed on all peripheral blood T, B and null cells as well as on thymocytes, 30% of bone marrow cells, eosinophils (stronger than on neutrophils) and epithelium of bronchus and salivary glands, but not on erythrocytes [1,2]. In the rat the MRC OX-45 antigen is expressed on all haematopoietic cells including erythrocytes and on endothelium [3]. All mouse thymocytes and lymph node cells as well as most macrophages and bone marrow cells express CD48, but erythrocytes do not express detectable amounts [4].

Structure

CD48 consists of an Ig V-like domain which lacks a disulphide bond and a Ig C2set domain which is linked to the lipid bilayer by a GPI anchor [5,6]. The site of GPI attachment is shown as Ser_{194} on the basis of homology with the rat sequence [7]. The sequence of CD48 matches particularly well with that of CD58 (LFA-3) and CD2 [6,7].

Function

In the rat system, anti-CD48 mAbs inhibit the MLR by activation of non-specific suppression by accessory cells, probably macrophages [8]. A recent study showed that GPI-anchored molecules of human and mouse T lymphocytes can associate with tyrosine kinases ($p56^{lck}$ in the case of human CD48) [9]. This interaction may be the mechanism for signal transduction by GPI-anchored molecules.

Comment

CD2 and CD58 (LFA-3) are linked to an ATPase gene in man and mouse. A similar linkage for the CD48 gene to another ATPase gene suggests that these linkage groups arose by duplication of a large chromosomal fragment [4].

Database accession numbers

	PIR	*SWISSPROT*	*EMBL/GENBANK*	*REFERENCE*
Human	A27236	P09326	X06341,M37766, M59904	6
Rat	S01299	P10252	X13016	7
Mouse	JL0143	P18181	X17501	5

Amino acid sequence of human CD48

```
MWSRGWDSCL ALELLLLPLS LLVTSI                                  -1
QGHLVHMTVV SGSNVTLNIS ESLPENYKQL TWFYTFDQKI VEWDSRKSKY        50
FESKFKGRVR LDPQSGALYI SKVQKEDNST YIMRVLKKTG NEQEWKIKLQ       100
VLDPVPKPVI KIEKIEDMDD NCYLKLSCVI PGESVNYTWY GDKRPFPKEL       150
QNSVLETTLM PHNYSRCYTC QVSNSVSSKN GTVCLSPPCT LARS             194
FGVEWIASWL VVTVPTILGL LLT                                    217
```

References

1 Vaughan, H.A. et al. (1983) Transplantation 36, 446–450.
2 Hadam, M.R. et al. (1989) Leukocyte Typing IV, 661–664.
3 Arvieux, J. et al. (1986) Mol. Immunol. 23, 983–990.
4 **Wong, Y.W. et al. (1990) J. Exp. Med. 171, 2115–2130.**
5 Staunton, D.E. and Thorley-Lawson, D.A. (1987) EMBO J. 6, 3695–3701.
6 Killeen, N. et al. (1988) EMBO J. 7, 3087–3091.
7 Staunton, D.E. et al. (1989) J. Exp. Med. 169, 1087–1099.
8 Arvieux, J. et al. (1986) Immunology 58, 337–342.
9 Stefanova, I et al. (1991) Science 254, 1016–1019.

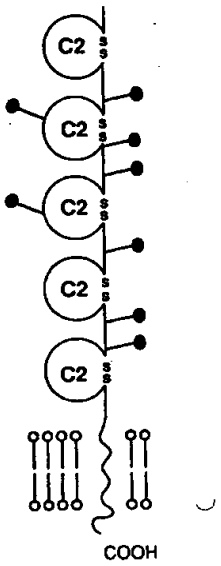
CD54 ICAM - 1

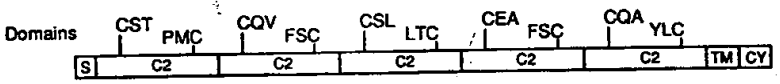
Molecular weights
Polypeptide 55402

SDS PAGE
reduced 85–110 kD

Carbohydrate
N-linked sites 8
O-linked unknown

Human gene location and size
19; 12 kb[11]





Tissue distribution

CD54 has a wide tissue distribution and is expressed on both haematopoietic and nonhaematopoietic cells [1,2]. Expression on resting peripheral blood leucocytes is weak but is upregulated on activated T and B lymphocytes and monocytes [1,2]. CD54 expression on endothelium and other nonhaematopoietic cells is upregulated by inflammatory mediators [1,2].

Structure

The extracellular domain is made up of five IgSF domains [1,2]. Within the IgSF, CD54 is most similar to ICAM-2 which has two IgSF domains that are closely related to CD54 domains 1 and 2 [2,3].

Function

CD54 mediates cell adhesion by binding to the integrins CD11a/CD18 (LFA-1) [1,2,3] and to CD11b/CD18 (Mac-1) [4]. This adhesion is not due to interaction with an epitope containing Arg-Gly-Asp (RGD) because this sequence is absent from CD54 [1,3]. The first NH_2-terminal domain of CD54 binds to CD11a/CD18 (LFA-1) [3] and the third to CD11b/CD18 [4]. The first domain of CD54 also binds to rhinovirus [3]. The interaction of CD54 with CD11a/CD18 enhances antigen-specific T cell activation [5]. CD54 has also been reported to bind to CD43 [6] and to *Plasmodium falciparum*- infected erythrocytes [9,10].

Database accession numbers

	PIR	*SWISSPROT*	*EMBL/GENBANK*	*REFERENCE*
Human	S00573	P05362	X06990	7
Mouse	S06015	P13597	X52264	8

Amino acid sequence of human CD54

```
MAPSSPRPAL PALLVLLGAL FPGPG                                    -1
NAQTSVSPSK VILPRGGSVL VTCSTSCDQP KLLGIETPLP KKELLLPGNN          50
RKVYELSNVQ EDSQPMCYSN CPDGQSTAKT FLTVYWTPER VELAPLPSWQ         100
PVGKNLTLRC QVEGGAPRAN LTVVLLRGEK ELKREPAVGE PAEVTTTVLV         150
RRDHHGANFS CRTELDLRPQ GLELFENTSA PYQLQTFVLP ATPPQLVSPR         200
VLEVDTQGTV VCSLDGLFPV SEAQVHLALG DQRLNPTVTY GNDSFSAKAS         250
VSVTAEDEGT QRLTCAVILG NQSQETLQTV TIYSFPAPNV ILTKPEVSEG         300
TEVTVKCEAH PRAKVTLNGV PAQPLGPRAQ LLLKATPEDN GRSFSCSATL         350
EVAGQLIHKN QTRELRVLYG PRLDERDCPG NWTWPENSQQ TPMCQAWGNP         400
LPELKCLKDG TFPLPIGESV TVTRDLEGTY LCRARSTQGE VTREVTVNVL         450
SPRYEIVIIT VVAAAVIMGT AGLSTYLYNR QRKIKKYRLQ QAQKGTPMKP         500
NTQATPP                                                        507
```


References

1 Dustin, M.L. and Springer, T.A. (1991) Annu. Rev. Immunol. 9, 27–66.
2 **Springer, T.A. (1990) Nature 346, 425–434.**
3 Staunton, D.E. et al. (1990) Cell 61, 243–254.
4 Diamond, M.S. et al. (1991) Cell 65, 961–971.
5 Altmann, D.M. et al. (1989) Nature 338, 512–514.
6 Rosenstein, Y. et al. (1991) Nature 354, 233–235.
7 Simmons, D. et al.(1988) Nature 331, 624–627.
8 Ballantyne, C.M. et al. (1989) Nucl. Acids Res. 17, 5853–5853.
9 Ockenhouse, C.F. et al. (1992) Cell 68, 63–69.
10 Berendt, A.R. et al. (1992) Cell 68, 71–81.
11 Voraberger, G. et al. (1991) J. Immunol. 147, 2777-2786.

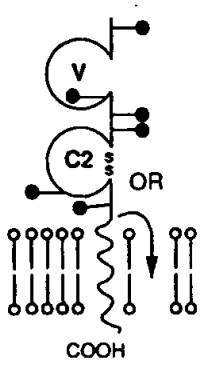
CD58 LFA-3

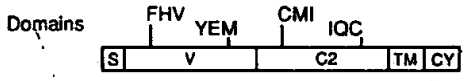
Molecular weights
Polypeptide: transmembrane 25339

SDS PAGE
reduced 55–70 kD
unreduced 55–70 kD

Carbohydrate
N-linked sites 6
O-linked unknown

Human gene location
1p13 [1]





Tissue distribution

CD58 has a wide tissue distribution on haematopoietic and various nonhaematopoietic cells including endothelium [2,3,4]. It is strongly expressed on macrophages and is present on germinal centre B cells [4], medullary thymocytes [4] and memory T lymphocytes [3]. Expression of CD58 on human cortical thymic epithelium is disputed [4]. The sheep CD58 homologue is expressed on erythrocytes and mediates the phenomenon of E rosetting, i.e. the rosettes formed spontaneously between sheep erythrocytes and human T cells.

Structure

The extracellular domain is similar to its ligand CD2 and also to CD48 [5]. It is made up of two IgSF domains with the NH_2-terminal one lacking the normally conserved disulphide bond [1,2,3,5]. Alternative splicing gives rise to a transmembrane form and a glycosyl-phosphatidylinositol anchored form [2,3]. The NH_2-terminus of the mature polypeptide has been established by protein sequencing [8].

Function

CD58 mediates cell adhesion through the binding of the extracellullar domain to the NH_2-terminal domain of CD2 [1,2,3]. Differences in function between the transmembrane and glycolipid-anchored form are unknown [2,3]. The affinity of CD58 for CD2 is in the order of 10^{-6} M [1,2]. Binding of CD58 to CD2 can enhance antigen-specific T cell activation [1,2]. A role for CD58 in signal transduction is suggested by the CD58-stimulated release of IL1 from thymic epithelial cells [6] or monocytes [7].

Comments

The genes for CD58 and CD2 are linked in the genome [1].

Database accession numbers

	PIR	SWISSPROT	EMBL/GENBANK	REFERENCE
Human	A28564	P19256	Y00636	8
Human (GPI)	S01269		X06296	9

Amino acid sequence of human CD58

```
MLRLLLALNL FPSIQVTGNK ILVKQSPM                                  -1
FSQQIYGVVY GNVTFHVPSN VPLKEVLWKK QKDKVAELEN SEFRAFSSFK           50
NRVYLDTVSG SLTIYNLTSS DEDEYEMESP NITDTMKFFL YVLESLPSPT          100
LTCALTNGSI EVQCMIPEHY NSHRGLIMYS WDCPMEQCKR NSTSIYFKME          150
NDLPQKIQCT LSNPLFNTTS SIILTTCIPS SGHSRHRYAL IPIPLAVITT          200
CIVLYMNGIL KCDRKPDRTN SN                                        222
```

GPI-anchored form (Ser_{180} is a potential site for GPI attachment)

```
SGHSRHRYAL IPIPLAVITT CIVLYMNVL                                 209
```

References

1 Moingeon, P. et al. (1989) Immunol. Rev. 111, 111–144.
2 Springer, T.A. et al. (1987) Annu. Rev. Immunol. 5, 223–252.
3 **Dustin, M.L. and Springer, T.A (1991) Annu. Rev. Immunol. 9, 27–66.**
4 Smith, M.E.F. and Thomas, J.A. (1990) J. Clin. Pathol. 43, 893–900.
5 Killeen, N. et al. (1988) EMBO J. 7, 3087–3091.
6 Le, P.T. et al. (1990) J. Immunol. 144, 4541–4547.
7 Webb, D.S.A. et al. (1990) Science 249, 1295–1297.
8 Wallner, B.P et al. (1987) J. Exp. Med. 166, 923–932.
9 Seed, B. (1987) Nature 329, 840–843.

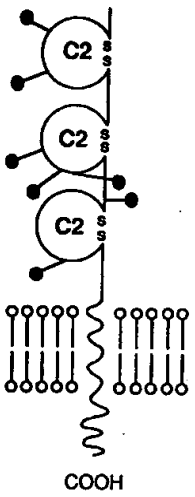
CD64 FcγRI

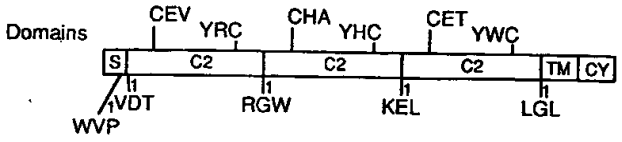
Molecular weights
Polypeptide 40833

SDS PAGE
reduced 72 kD

Carbohydrate
N-linked sites 7
O-linked unknown

Human gene location and size
1q, 9.4 kb [1]





Tissue distribution

Human CD64 is expressed on macrophages, monocytes and interferon γ induced neutrophils. Interferon γ treatment strongly upregulates expression on monocytes. Mouse CD64 is expressed on macrophages [2,3].

Structure

The extracellular region consists of three C2-set IgSF domains. The first two domains are about 50% identical to the extracellular domains of CDw32 (FcγRII) and CD16 (FcγRIII) and altogether the Fc receptors form a closely related subgroup within the IgSF [4]. The third and less closely related domain has been proposed to be responsible for the high affinity binding of CD64 to IgG [2,4]. Two alleles of human CD64 have been described [2]. A third form with a different cytoplasmic sequence may be the result of a cloning artifact [2,4].

Function

CD64 is a high affinity receptor for IgG and the only IgG receptor which binds to monomeric ligand [2]. The FcγRI binding site in IgG_1 has been mapped to the CH2 domain [5]. FcγRI plays a central role in antibody-dependent cytotoxicity and clearance of immune complexes [3].

Database accession numbers

	PIR	*SWISSPROT*	*EMBL/GENBANK*	*REFERENCE*
Human	S03018	P12314	X14356	4
	S03019	P12315	X14355	4
Mouse			M31314	6

Amino acid sequence of human CD64

SWISSPROT: P12314

```
MWFLTTLLLW VPVDG                                                -1
          T*                                          V*
QVDTTKAVIS LQPPWVSVFQ EETVTLHCEV LHLPGSSSTQ WFLNGTATQT   50
STPSYRITSA SVNDSGEYRC QRGLSGRSDP IQLEIHRGWL LLQVSSRVFT  100
EGEPLALRCH AWKDKLVYNV LYYRNGKAFK FFHWNSNLTI LKTNISHNGT  150
YHCSGMGKHR YTSAGISVTV KELFPAPVLN ASVTSPLLEG NLVTLSCETK  200
LLLQRPGLQL YFSFYMGSKT LRGRNTSSEY QILTARREDS GLYWCEAATE  250
DGNVLKRSPE LELQVLGLQL PTPVWFHVLF YLAVGIMFLV NTVLWVTIRK  300
ELKRKKKWDL EISLDSGHEK KVTSSLQEDR HLEEELKCQE QKEEQLQEGV  350
HRKEPQGAT                                               359
```

* Allelic residues in human CD64.

SWISSPROT: P12315. Possible cloning artifact.

```
MWFLTTLLLW VPVDG                                                -1
QVDTTKAVIT LQPPWVSVFQ EETVTLHCEV LHLPGSSSTQ WFLNGTATQT   50
STPSYRITSA SVNDSGEYRC QRGLSGRSDP IQLEIHRGWL LLQVSSRVFT  100
EGEPLALRCH AWKDKLVYNV LYYRNGKAFK FFHWNSNLTI LKTNISHNGT  150
YHCSGMGKHR YTSAGISVTV KELFPAPVLN ASVTSPLLEG NLVTLSCETK  200
LLLQRPGLQL YFSFYMGSKT LRGRNTSSEY QILTARREDS GLYWCEAATE  250
DGNVLKRSPE LELQVLGLQL PTPVWFHVLF YLAVGIMFLV NTVLWVTIRK  300
ELKRKKKWDL EISLDSGGQA LEAPTQGCA                         329
```

References

1 van de Winkel, J.G.J. et al. (1991) J. Biol. Chem. 266, 13449–13455.
2 **Ravetch, J.V. and Kinet, J.P. (1991) Annu. Rev. Immunol. 9, 457–492.**
3 **Unkeless, J.C. et al. (1988) Annu. Rev. Immunol. 6, 251–281.**
4 Allen, J.M. and Seed, B. (1989) Science 243, 378–380.
5 Chappel, M.S. et al. (1991) Proc. Natl Acad. Sci. USA 88, 9036–9040.
6 Sears, D.W. et al. (1990) J. Immunol. 144, 371–378